PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Review of Quality Assessment Tools for the Evaluation of Pharmacoepidemiologic Safety Studies
AUTHORS	Neyarapally, George ; Hammad, Tarek; Pinheiro, Simone; Iyasu, Solomon

VERSION 1 - REVIEW

REVIEWER	Ivo Abraham Professor Center for Health Outcomes and PharmacoEconomic Research University of Arizona Tucson, AZ, USA Chief Scientist Matrix45 Tucson, AZ, USA No competing interests to disclose.
REVIEW RETURNED	01-Jun-2012

THE STUDY	As a general comment, since this is an evaluation of tools, questions above referring to participants and patients are answered with this substitution in mind. This paper summarizes the findings of a systematic inventory/evaluation/review of tools. This is an important exercise as the field is indeed replete with tools, most of which with limited if any validation. Thus the authors' efforts to plow through a number of tools and evaluate them systematically could potentially yield important information. Summarized, this paper has merit in the exercise it reports but the implications of the findings may require more thought. However, one can question whether this is sufficient, especially the way the results are reported. The Results section is rather nondescript and instead the reader is referred to the accompanying set of tables/figures. Here we find aggregate descriptive results of the extent to which the tools included in the review meet or do not meet a (well-developed) set of (credible) criteria. This may very well give us an impression of the poor state-of-the-art in the domain of interest. However, the "so what?" question of the research remains unanswered - or, in a revision, should be specified more clearly. What is the relevance of knowing that tools in this domain tend to be of limited quality and relevance? How does this help the field move forward? Specifically, isn't there merit to classify the tools in categories of quality so that we can at least get a sense of which tools might be appropriate and useful?

REVIEWER	Carlo Marra Associate Professor, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, CANADA
REVIEW RETURNED	12-Jun-2012

GENERAL COMMENTS	This is a well-written, thoughtful paper. The only concern that I had was in the selection of the criteria that were used to evaluate the tools. While the criteria appear to have face validity, a more systematic process to establish the criteria (modified delphi or something similar) would have been useful. Perhaps the authors could comment on this.

REVIEWER	Dr Alex McMahon Reader in Epidemiology, University of Glasgow Dental School.
REVIEW RETURNED	22-Jun-2012

GENERAL COMMENTS	This paper aims to address the issue of guidelines in pharmacoepidemiology research and publication. There is a useful paper to be had out of this area of research, but leading to a different line of thought than that presented in the manuscript. (1) My first major comment is that the paper needs a restructure. There are two types of guidelines: reporting guidelines (analogous to CONSORT) and quality assessment guidelines (analogous to Jadad). There are two different types of epidemiology: pharmacoepidemiology and the others. Other types of epidemiology could perhaps do with bespoke guidelines as well (eg genetic epidemiology). However, pharmacoepidemiology is sufficiently different to need special handling. These four combinations need separating out. (2) The main point of this paper is apparently to identify quality assessment guidelines for pharmacoepidemiology studies. You don't need a literature review to discover that there is no such thing. If the paper was partitioned as mentioned above then this would be more tightly focussed. (3) (cf throughout and Page 10) Another major design issue is the contrast between intended and unintended effects (eg efficacy and safety). There are huge differences in quality between these two types of study in pharmacoepidemiology. Comparative effectiveness studies are much more controversial (similar to the discredited lifestyle risk factor cohort studies that are published almost weekly, and are fatally cofounded by economic status rather than by confounding by indication). However, pharmacoepidemiology is the best available study design for unintended drug toxicity problems (and is even more useful than RCTs in this regard). A first go at pharmacoepidemiology guidelines should concentrate on the unintended effects of drugs. (4) The paper could then call for such a guideline to be created, if the authors think this is important. In my opinion there is no general consensus on study design in this field.

	(5) Note that the Cochrane review for clinical trials picked SIGN50 over the Jadad scale, which is interesting. Methodological domains are emphasised. A useful pharmacoepidemiology guideline would need to make decisions on the features of study design that matter most. Perhaps the authors should start the ball rolling with some suggestions, remembering that pharmacoepidemiology designs are quite different from other studies in epidemiology. (6) Unfortunately there are only a small number of methodologists worldwide who could contribute towards this type of work. David Sackett once humorously noted that he would only trust half a dozen people to design a case-control study. I myself used to dabble in the pros and cons of study design, so I should declare an interest: McMahon AD, MacDonald TM. Design issues in drug epidemiology. British Journal of Clinical Pharmacology 2000; 50; 5: 419-425. McMahon AD. Observation and experiment with the efficacy of drugs: a warning example using a cohort of NSAID and ulcer healing drug users. American Journal of Epidemiology 2001; 154; 6: 557-562. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. Pharmacoepidemiology and Drug Safety 2003; 12; 7: 551-558. Any serious attempt to create a quality guideline would have to assemble a group of published methodologists. Without naming names it would be reasonably easy to assemble opinion leaders in study design. After all, one of the reasons that CONSORT was so widely applied and is genuinely respected is that the right people were 'in the room', as the saying goes. It should be recognised that although a consensus is possible, it would be very difficult. Disagreements would have to be settled at an intellectual level, using the best available evidence, with occasional opinions having to be overruled. (7) The paper should include some opinion on what are the most important areas of study design. Non-experts need guidelines on (for example); new user designs; clean cohorts restricted to subjects without strong indications for a single drug, rather than a comparator, and without contraindications; what to do with drug switchers and combination users; exposure timing windows; and whether or not only cohort studies should be used, leaving case-control samples for more complicated scenarios involving some field work. (8) The authors note on page 1 that observational studies have fewer exclusion criteria than RCTs, this is arguably part of the problem. Observational studies should be more restricted to control confounding, and RCTs should have as few restrictions as possible. I've always thought that this may lead to a rough convergence of the two paradigms with regard to who enters the study. (9) There are some fairly well known papers that together could be considered as a sort of guidebook to pharmacoepidemiology study design. In my opinion these two are seminal works: Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. J Clin Epidemiol 1989; 42: 325-331. Guess HA. Behaviour of the exposure odds ratio in a case-control study when the hazard function is not constant over time. J Clin

	Epidemiol 1989; 42: 1179-1184. The recent paper by Schneeweiss is also very useful, and provides a more recent assessment of the 'state of play' in drug safety monitoring. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf. 2010 August ; 19(8): 858-68. (10) I found the 100-odd uses of the word 'tool' really irritating, it does not need to be used so often. The authors should try and reduce the 'tool' count to under a dozen!

VERSION 1 – AUTHOR RESPONSE

Reviewer: Ivo Abraham

Professor

Center for Health Outcomes and PharmacoEconomic Research

University of Arizona

Tucson, AZ, USA

Chief Scientist

Matrix45

Tucson, AZ, USA

No competing interests to disclose.

As a general comment, since this is an evaluation of tools, questions above referring to participants and patients are answered with this substitution in mind.

This paper summarizes the findings of a systematic inventory/evaluation/review of tools. This is an important exercise as the field is indeed replete with tools, most of which with limited if any validation. Thus the authors' efforts to plow through a number of tools and evaluate them systematically could potentially yield important information.

Summarized, this paper has merit in the exercise it reports but the implications of the findings may require more thought.

However, one can question whether this is sufficient, especially the way the results are reported. The Results section is rather nondescript and instead the reader is referred to the accompanying set of tables/figures. Here we find aggregate descriptive results of the extent to which the tools included in the review meet or do not meet a (well-developed) set of (credible) criteria.

This may very well give us an impression of the poor state-of-the-art in the domain of interest. However, the "so what?" question of the research remains unanswered - or, in a revision, should be specified more clearly. What is the relevance of knowing that tools in this domain tend to be of limited quality and relevance? How does this help the field move forward? Specifically, isn't there merit to classify the tools in categories of quality so that we can at least get a sense of which tools might be appropriate and useful?

Response: We thank the reviewer for stating that our criteria are well-developed and credible. As we mentioned in the paper, the reporting elements and quality assessment attributes constitute key considerations to facilitate a robust, consistent and transparent evaluation of pharmacoepidemiologic safety studies. However, we did not create a validated tool as we aimed to first systematically evaluate the available tools, establish that there is a critical gap in our tool kit, and then encourage timely discussion in the research community about the potential value of developing such a tool. Our findings demonstrate that (1) tools reviewed were not designed to evaluate pharmacoepidemiologic safety studies and (2) the tools did not comprehensively address the reporting elements and quality assessment attributes in our preliminary assessment framework, which suggests that there is a need for the development of a validated tool. As we did not find any tools adequate for the robust evaluation of pharmacoepidemiologic safety studies, we did not classify tools in different categories of quality as the reviewer suggests. If and when there is agreement on the need for a validated tool, relevant stakeholders could start the process of developing a comprehensive assessment tool specifically designed for pharmacoepidemiologic safety studies, which would first involve reviewing relevant articles, guidelines and guidance documents, identifying the key domains, and then developing reporting elements and assessment attributes.

With respect to the comment regarding the non-descript nature of the results, we added specific, relevant results to the text as follows **[see pages 7 - 8]:**

“Representation of a priori assessment domains and elements within tools

The proportion of reviewed tools that included reporting elements (RE) and quality assessment attributes (QAA) according to each a-priori defined domain within the framework is shown in Figure 1. Table 1 depicts the detailed results of our review of the domains, elements and attributes. We highlighted the representation of select RE and QAA under each domain that may have important implications for the assessment of a pharmacoepidemiologic safety study. RE and QAA related to research aims were addressed in 69% (42/61) and 34% (21/61) of the tools, respectively. Regarding the domain assessing study population and data sources, 84% (51/61) of the tools included RE and 57% (35/61) included QAA (Table 1).

61% (37/61) of the tools included RE and 31% (19/61) included QAA under the exposure definition and ascertainment domain (Table 1). With respect to outcome definition and ascertainment

domain, 69% (42/61) of the tools included RE and 36% (22/61) included QAA (Table 1). Out of the 61 reviewed tools, 85% (52/61) and 49% (30/61) included RE and QAA under the analytic approach domain (Table 1). Under the results domain, only 36% (22/61) and 7% (4/61) of tools included RE and QAA respectively (Table 1).

Of the 61 reviewed tools, 36% (22/61) and 20% (12/61) of tools included RE and QAA under the discussion and interpretation domain (Table 1). 7% (4/61) of the tools addressed the description of the study team (RE) and the independence of team and funding sources (QAA)."

Regarding the "so what" question with respect to our findings, it is important to know that existing tools do not address key elements and attributes of study quality. This highlights the potential utility of developing an assessment tool to evaluate these studies to ensure transparent, objective and consistent evaluations, although it would not obviate the need for clinical, epidemiological, and statistical expertise in the evaluation process. Consideration of evidence from poorly designed/conducted studies (e.g., poor exposure or outcome ascertainment) present challenges for regulators, clinicians and other stakeholders. For example, if a poorly designed study finds no association between a drug and harm of interest, this may be falsely reassuring in light of all available evidence and result in misinformed regulatory and clinical decision making, which would lead to patients staying on the drug, thus leading to adverse patient health outcomes. Conversely, a poorly designed/conducted study may find an association between a drug and a harm of interest which may erroneously garner more weight in a safety-related regulatory or clinical decision making and potentially resulting in patients being switched to other drugs, which might have a less favorable benefit-risk profile. As sometimes there are multiple pharmacoepidemiologic safety studies that influence regulatory or clinical decision making, a validated quality assessment tool could ensure that the quality of studies are consistently reviewed and well designed/conducted studies receive the appropriate weight in regulatory and clinical decision making across different drug safety issues.

Finally, the Institute of Medicine report recently pointed out that often times there are significant disagreements regarding scientific evidence used for regulatory decision making, including pharmacoepidemiologic safety studies[1] (this sentence was added to the manuscript) **[see pages 2-3].** This review stimulates discussion about the potential utility in developing an assessment tool to facilitate transparency and consistency in the evaluation of these studies, which would mitigate the uncertainty about how this scientific evidence is used across different safety issues, especially in light of potential conflicting results from studies and scientific disagreements about the evidence stemming from such studies.

Reviewer: Carlo Marra

Associate Professor, Faculty of Pharmaceutical Sciences,

University of British Columbia, Vancouver, CANADA

This is a well-written, thoughtful paper. The only concern that I had was in the selection of the criteria that were used to evaluate the tools. While the criteria appear to have face validity, a more systematic

[1] Institute of Medicine. Ethical and Scientific Issues in Studying the Safety of Approved Drugs. 2012.

process to establish the criteria (modified delphi or something similar) would have been useful. Perhaps the authors could comment on this.

Response: We thank the reviewer for his comments. As we mentioned in the paper, the reporting elements and quality assessment attributes constitute key considerations for the evaluation of pharmacoepidemiologic safety studies and do not constitute all pertinent elements and attributes. The aim of our work was to stimulate discussion in the research community about the potential utility of a (validated) tool to improve the consistency and transparency of the evaluation of these pharmacoepidemiology studies. In line with this objective, we developed a preliminary assessment framework to evaluate the utility of available assessment tools for the assessment of pharmacoepidemiologic safety studies. If a validated tool is developed in the future, perhaps by a group of expert methodologists, we agree that a systematic approach should be used in its development. In addition, if a tool is developed, the expert group should engage and obtain input from a variety of stakeholders such as regulators, healthcare professionals, epidemiologists, patients, non-governmental organizations, clinical guideline developers, payors, and others, to ensure buy in and develop a tool, i.e., “develop the criteria.”

Reviewer: Dr Alex McMahon

Reader in Epidemiology,

University of Glasgow Dental School.

This paper aims to address the issue of guidelines in pharmacoepidemiology research and publication. There is a useful paper to be had out of this area of research, but leading to a different line of thought than that presented in the manuscript.

(1) My first major comment is that the paper needs a restructure. There are two types of guidelines: reporting guidelines (analogous to CONSORT) and quality assessment guidelines (analogous to Jadad). There are two different types of epidemiology: pharmacoepidemiology and the others. Other types of epidemiology could perhaps do with bespoke guidelines as well (eg genetic epidemiology). However, pharmacoepidemiology is sufficiently different to need special handling. These four combinations need separating out.

Response: We thank the reviewer for the detailed comments on our paper. We think the reviewer brings up a good point about differences between reporting and quality assessment as well as between pharmacoepidemiology and other fields of epidemiology. We added the following language in tracked changes to emphasize this point which we discussed generally in the paper:

“Although many checklists and scales for the assessment of epidemiologic studies exist, most are not specifically designed to evaluate pharmacoepidemiologic safety studies. **Importantly, although the principles of epidemiology apply across different fields, there are unique considerations and challenges in the design, conduct and evaluation of epidemiologic studies of unintended drug harms that warrant consideration of developing a specific validated assessment tool (e.g.,**

confounding by indication is an important challenge that is unique to epidemiologic studies of drug effects).” Recent articles have suggested the need to develop tools for assessing the quality of these studies. A recent publication found that systematic reviewers and meta-analysts are misusing reporting tools like STROBE due to the dearth of validated assessment instruments.” [**see page 4**]

“Within each domain we listed critical elements that need to be considered for assessing the validity and interpretation of findings from such studies. We made a distinction between the reporting elements (RE) and quality assessment attributes (QAA) for each domain. This is an important distinction as some guidelines are strictly developed to discern and evaluate reporting whereas other tools are developed to evaluate quality, which requires assessment of reporting…” **[see page 5]**

(2) The main point of this paper is apparently to identify quality assessment guidelines for pharmacoepidemiology studies. You don’t need a literature review to discover that there is no such thing. If the paper was partitioned as mentioned above then this would be more tightly focussed.

Response: We performed a comprehensive literature review to determine if there were tools designed to evaluate pharmacoepidemiologic safety studies and to determine if existing tools, even if not specifically designed to evaluate these studies, could be used for their evaluation based on domains, reporting elements and assessment attributes.

(3) (cf throughout and Page 10) Another major design issue is the contrast between intended and unintended effects (eg efficacy and safety). There are huge differences in quality between these two types of study in pharmacoepidemiology. Comparative effectiveness studies are much more controversial (similar to the discredited lifestyle risk factor cohort studies that are published almost weekly, and are fatally cofounded by economic status rather than by confounding by indication). However, pharmacoepidemiology is the best available study design for unintended drug toxicity problems (and is even more useful than RCTs in this regard). A first go at pharmacoepidemiology guidelines should concentrate on the unintended effects of drugs.

Response: We agree with the reviewer that there are important differences between (1) observational studies designed to evaluate drug benefits and (2) observational studies designed to evaluate drug harms. Regulators and other stakeholders (e.g., clinical guideline developers) evaluate observational pharmacoepidemiologic studies of purported associations between drugs and harms. Thus, our preliminary assessment framework focuses on the evaluation of pharmacoepidemiologic safety studies (studies of unintended drug harms), which is consistent with the reviewer’s support of the development of guidelines concentrating on the unintended effects of drugs. In addition, in the paper we pointed out that “critical assessment elements of pharmacoepidemiologic studies focused on effectiveness may be different than those focused on safety…” [**see page 11**]

(4) The paper could then call for such a guideline to be created, if the authors think this is important. In my opinion there is no general consensus on study design in this field.

Response: The purpose of our evaluation of tools with a preliminary assessment framework was to stimulate a discussion in the research community about the potential utility of quality assessment tools for the evaluation of pharmacoepidemiologic studies of drug harms. In the discussion section, we stated that "the development of an assessment tool based on expert input may facilitate consistent, evidence-based quality assessment of such studies and the subsequent determination of their value based on evaluating the impact of bias on the robustness of a study results, and the interpretation of its findings, within the context of the specific drug safety issue." **[see page 12**] In addition, regulators and other decision makers should provide for public input from stakeholders on the need to develop a tool.

(5) Note that the Cochrane review for clinical trials picked SIGN50 over the Jadad scale, which is interesting. Methodological domains are emphasised. A useful pharmacoepidemiology guideline would need to make decisions on the features of study design that matter most. Perhaps the authors should start the ball rolling with some suggestions, remembering that pharmacoepidemiology designs are quite different from other studies in epidemiology.

Response: The purpose of this review was to stimulate a discussion in the research community about the potential utility of quality assessment tools for the evaluation of pharmacoepidemiologic studies of drug harms. Making recommendations regarding features of study designs that matter most is beyond the scope of this review. However, there are multiple published documents that provide guidelines on pharmacoepidemiologic studies, including, for example, FDA's 2011 draft guidance,[2] ISPE's guidelines[3], ENCePP's methods guide.[4] Nonetheless, our preliminary assessment framework does include important attributes we consider necessary to fully evaluate a study. In addition, we pointed out in the paper that "Although we did not address weighing of importance of different domains and elements based on their relative impact on study contribution to the available streams of evidence, this may be an important consideration in the formulation of an assessment tool…."[92] **[see page 11]** Thus, this an important research gap that should be addressed if a validated tool is developed.

(6) Unfortunately there are only a small number of methodologists worldwide who could contribute towards this type of work. David Sackett once humorously noted that he would only trust half a dozen people to design a case-control study. I myself used to dabble in the pros and cons of study design, so I should declare an interest:

McMahon AD, MacDonald TM. Design issues in drug epidemiology. British Journal of Clinical Pharmacology 2000; 50; 5: 419-425.

[2] FDA Draft Guidance: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets; 2011.
[3] ISPE guidelines. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiology and Drug Safety 2008;17:200-208.
[4] The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology.
http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE.pdf.

McMahon AD. Observation and experiment with the efficacy of drugs: a warning example using a cohort of NSAID and ulcer healing drug users. American Journal of Epidemiology 2001; 154; 6: 557-562.

McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. Pharmacoepidemiology and Drug Safety 2003; 12; 7: 551-558.

Any serious attempt to create a quality guideline would have to assemble a group of published methodologists. Without naming names it would be reasonably easy to assemble opinion leaders in study design. After all, one of the reasons that CONSORT was so widely applied and is genuinely respected is that the right people were 'in the room', as the saying goes. It should be recognised that although a consensus is possible, it would be very difficult. Disagreements would have to be settled at an intellectual level, using the best available evidence, with occasional opinions having to be overruled.

Response: We agree that the creation of a quality assessment tool, if pursued, would ideally be developed by a group of expert methodologists. In the paper, we stated, "The creation of this tool could be led by an independent expert or academic group, perhaps with input from regulatory agencies." We have amended this sentence as follows to emphasize the need for methods expertise: "The creation of this instrument could be led by an independent group of expert methodologists, perhaps with input from multiple stakeholders, including regulators.and professional organizations." **[see page 11]**

(7) The paper should include some opinion on what are the most important areas of study design. Non-experts need guidelines on (for example); new user designs; clean cohorts restricted to subjects without strong indications for a single drug, rather than a comparator, and without contraindications; what to do with drug switchers and combination users; exposure timing windows; and whether or not only cohort studies should be used, leaving case-control samples for more complicated scenarios involving some field work.

Response: Although we agree that it is important to identify the most important areas of study design for purposes of evaluation, the purpose of this review was to stimulate a discussion in the research community about the potential utility of quality assessment tools for the evaluation of pharmacoepidemiologic studies of drug harms. Making specific recommendations regarding areas of study designs is beyond the scope of this review.

(8) The authors note on page 1 that observational studies have fewer exclusion criteria than RCTs, this is arguably part of the problem. Observational studies should be more restricted to control confounding, and RCTs should have as few restrictions as possible. I've always thought that this may lead to a rough convergence of the two paradigms with regard to who enters the study.

A discussion of the issue of "restriction" in study design is beyond the scope of the current paper but we sincerely thank the reviewer for the thoughtful comments.

(9) There are some fairly well known papers that together could be considered as a sort of guidebook to pharmacoepidemiology study design. In my opinion these two are seminal works:

Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. J Clin Epidemiol 1989; 42: 325-331.

Guess HA. Behaviour of the exposure odds ratio in a case-control study when the hazard function is not constant over time. J Clin Epidemiol 1989; 42: 1179-1184.

The recent paper by Schneeweiss is also very useful, and provides a more recent assessment of the 'state of play' in drug safety monitoring.

Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf. 2010 August ; 19(8): 858-68.

Response: We thank the reviewer for providing these important references and added these articles to the reference list of the paper to serve as a resource to readers as follows.

"If after further consideration and discussions with stakeholders development of a tool to evaluate epidemiologic data for drug safety is pursued, it would be necessary to first determine the scope of the assessment tool as well as steps for its comprehensive validation. Further, relevant aspects of the design and analysis of pharmacoepidemiology studies should be considered (we refer the reader to some helpful references [cite 98, 99, and 100]). Importantly, such a tool would be intended to complement, and not replace, expert clinical, methodological, and statistical expertise necessary to complete a robust evaluation and determination of the contribution of a specific pharmacoepidemiologic safety study to the available evidence for regulatory decision making." **[see page 13]**

(10) I found the 100-odd uses of the word 'tool' really irritating, it does not need to be used so often. The authors should try and reduce the 'tool' count to under a dozen!

Response: We reduced the use of the word "tool" throughout the document (please refer to the tracked changes).

VERSION 2 – REVIEW

REVIEWER	Alex McMahon no competing interests
REVIEW RETURNED	17-Aug-2012

- The reviewer completed the checklist but made no further comments.